UNCLASSIFIED

AD 220545

FOR MICRO-CARD CONTROL ONLY

1 OF 1

Code 25

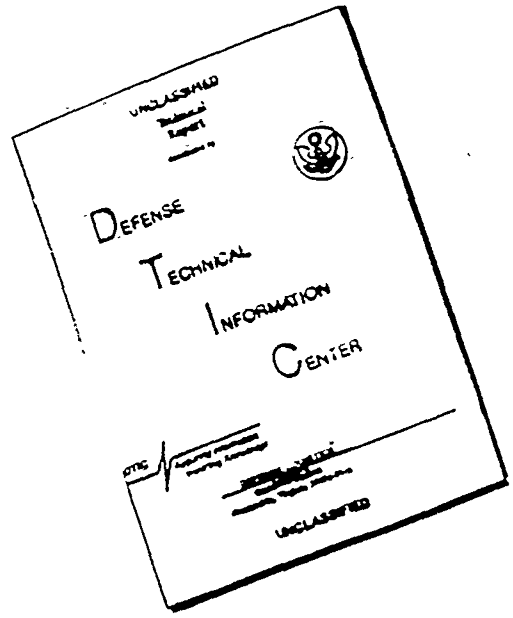
Reproduced by

Armed Services Technical Information Agency

ARLINGTON HALL STATION; ARLINGTON 12 VIRGINIA

UNCLASSIFIED

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

"NOTICE: When Government or other drawings, specifications
other data are used for any purpose other than in connection
a definitely related Government procurement operation, the U.S
Government thereby incurs no responsibility, nor any obligati
whatsoever; and the fact that the Government may have form
furnished, or in any way supplied the said drawings, specificati
or other data is not to be regarded by implication or otherwis
in any manner licensing the holder or any other person or co
tion, or conveying any rights or permission to manufacture, us
sell any patented invention that may in any way be related th

FILE COPY

Return to

ASTIA

ARLINGTON HALL STATION

ARLINGTON 12, VIRGINIA

Attn: TISSS

ONE-SIDED INEQUALITIES OF THE CHEBYSHEV TYPE

by

ALBERT W. MARSHALL AND INGRAM OLKIN

TECHNICAL REPORT NO. 48

August 6, 1959

PREPARED UNDER CONTRACT N6onr-25140

(NR 342-002)

OFFICE OF NAVAL RESEARCH

This work was sponsored by the Army, Navy, and Air Force through the Joint Services Advisory Committee for Research Groups in Applied Mathematics and Statistics by Contract No. N6onr-25140 (NR 342-002).

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

APPLIED MATHEMATICS AND STATISTICS LABORATORY

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

ASTIA
AUG 25 1959
TIPDR

PART I

Page

A one-sided inequality of the Chebyshev type,

Albert W. Marshall and Ingram Olkin 1

PART II

Page

A one-sided analog of Kolmogorov's inequality,

Albert W. Marshall 12

PREFACE

In many applications one is frequently confronted with the problem of obtaining a bound for the probability that a random vector falls in a certain region. Alternatively, the bound may be specified and one may want to define the limits of the region. This report is concerned with the theoretical development of two such bounds, and we now give several examples which illustrate the use of the results.

1. MULTIPLE DECISION PROCEDURES; SLIPPAGE PROBLEMS.

Several authors have considered these problems, e.g., Bechhofer, Dunnett, and Sobel; Paulson; Karlin and Truax. For this situation we have a random vector $t = (t_1, \ldots, t_k)$, where each t_i has a student's t-distribution. However, the t_i are correlated because the estimate of variance is the same in each t_i. In choosing the population with the largest mean or in deciding whether the mean in one of several populations has slipped, the relevant statistic is $\max_{1 \le j \le k} t_j$. Thus one requires information about $P\{\max t_j \ge c\} \le \alpha$. In this case one may desire the value of α if c is given, or the value of c if α is given.

Since the joint distribution of the t_i is known, one might hope to obtain more accurate bounds for $\max t_i$. However the distribution of $\max t_i$ is not easily determined so that Chebyshev type bounds can be used as an approximation.

Some typical problems where this arises are:

(i) The determination of which of several types of alloys will produce the highest mean tensile strength,

(ii) the determination of which of several drugs produces the greatest mean effect,

(iii) the determination of which of several lots of projectiles will produce the longest mean range.

2. ANALYSIS OF VARIANCE.

A parallel situation to 1 prevails in some types of analysis of variance problems. Here we have a random vector $F = (F_1, \ldots, F_k)$, where each F_i has an F-distribution, but the F_i are correlated. Again the relevant statistic is $\max_{1 \leq j \leq k} F_j$. Some tables for small values of k have been given by Ramachandran.

3. PHYCHOLOGICAL TESTING.

In certain testing problems there may be a number of tests which in some sense increase in difficulty. For example, an ordering of difficulty might be addition, multiplication, calculus. Following a model of Guttman, the score on test one is given by y_1, the score on test two by $y_1 + y_2$, and the score on test j by $y_1 + \cdots + y_j$. Thus the increase in difficulty from test j to test $j+1$ is represented by y_{j+1}. Given various assumptions on the y's, one is interested in some personnel assignment problems in finding a bound for the $P\left\{ \max_{1 \leq j \leq k} (y_1 + \ldots + y_j) \geq c \right\}$.

A ONE-SIDED INEQUALITY OF THE CHEBYSHEV TYPE

by

Albert W. Marshall and Ingram Olkin[1/]

0. Summary: Let $x = (x_1, \cdots, x_k)$ be a random vector with $Ex_i = 0$ and either (i) $Ex_i^2 = \sigma^2$, $Ex_i x_j = \sigma^2 \rho \ (i \neq j)$, $i, j = 1, \cdots, k$, or (ii) $Ex_i^2 = \sigma_i^2$, $i = 1, \cdots, k$. In each case a sharp upper bound (only involving the specified moments) for $P(x_1 \geq 1 \text{ or } \cdots \text{ or } x_k \geq 1)$ is obtained. A formulation for more general one-sided bounds is also discussed.

1. Introduction: Let x be a random variable with mean 0 and variance σ^2. Then according to Chebyshev's inequality,

$$(1.1) \qquad P(|x| \geq 1) \leq \sigma^2 ;$$

the corresponding one-sided inequality

$$(1.2) \qquad P(x \geq 1) \leq \sigma^2/(\sigma^2 + 1)$$

is also known (see e.g. [2, p.198]). Both inequalities are sharp.

A generalization of inequality (1.1) was obtained by Olkin and Pratt [1] for $P(|x_1| \geq 1 \text{ or } \cdots |x_k| \geq 1)$ in the case $Ex_i = 0$, $Ex_i^2 = \sigma^2$, $Ex_i x_j = \sigma^2 \rho \ (i \neq j)$, $i, j = 1, \cdots, k$. We are concerned here with the corresponding bound for $P(x \epsilon T) = P(x_1 \geq 1 \text{ or } \cdots \text{ or } x_k \geq 1)$.

1/ On leave of absence from Michigan State University. This research was sponsored in part by the Office of Ordnance Research.

Consider a real-valued function $f(x) \equiv f(x_1,\cdots,x_k)$ such that

$$(1.3) \qquad f(x) \geq 0 \quad \text{for all} \quad x \,, \quad f(x) \geq 1 \quad \text{for} \quad x \in T \,;$$

then

$$(1.4) \qquad E\, f(x) \geq \int_{\{x \in T\}} f(x)dP \geq P\{x \in T\} \,.$$

Since we wish the bound to be a function of the covariance matrix $\Sigma = (\sigma_{ij})$, we choose

$$(1.5) \qquad f(x) = (x-a)\, A\, (x-a)' \,,$$

where $a = (a_1,\cdots,a_k)$, $A = (a_{ij}) : k \times k$. A "best" bound is one which minimizes

$$(1.6) \qquad E\, f(x) = \operatorname{tr} \, A(\Sigma + a'a) \,,$$

subject to (1.3) . Let $z = (x-a)\, D_{1-a}^{-1}$, where $D_{1-a} = \operatorname{diag}\,(1-a_1,\cdots,1-a_k)$ and let $A^* = D_{1-a}\, A\, D_{1-a}$. Since $f(a) = 0$, (1.3) implies that $a \notin T$, i.e. $a_i < 1$, $i = 1,\cdots,k$. Thus (1.3) becomes

$$(1.7) \qquad z A^* z' \geq 0 \quad \text{for all} \quad z, \quad z A^* z' \geq 1 \quad \text{if some} \quad z_j \geq 1 \,.$$

By the results of [1] the bound (1.6) will be minimized by a positive

definite matrix A^* (or in turn A) for which $B = (A^*)^{-1}$ has ones on the main diagonal. Thus the problem has been transformed to that of minimizing

$$\text{(1.8)} \qquad \operatorname{tr} A(\Sigma + a'a) = \operatorname{tr} B^{-1} D_{1-a}^{-1} (\Sigma + a'a) D_{1-a}^{-1}$$

for $a_i < 1$ and $B = (b_{ij})$ positive definite with $b_{ii} = 1$, $i = 1,\cdots,k$. Any such a and B will yield a valid bound which in general will not be sharp.

The problem of finding a sharp bound has not been solved for general covariance matrices even when $k = 2$. We give a solution for the case that $\sigma_{ii} = \sigma^2$, $\sigma_{ij} = \sigma^2\rho$ $(i \neq j)$, and an example which attains equality in Sections 2 and 3. In Section 4 we give a sharp bound which only involves the variances.

We will frequently encounter a class $\mathcal{D}$ of positive definite matrices $\Delta = (\delta_{ij})$ with $\delta_{ii} = 1$, $\delta_{ij} = \delta (i \neq j)$. Then Δ can be written as

$$\text{(1.9)} \qquad \Delta = (1-\delta) I + \delta e'e ,$$

where $e = (1,\cdots,1)$. Such a matrix may be transformed to diagonal form by an orthogonal matrix Γ , whose first row is $e/\sqrt{k}$, i.e.

$$\text{(1.10)} \qquad \begin{aligned} \Gamma\Delta\Gamma' &= (1-\delta) I + \delta \Gamma e'e\Gamma \\ &= (1-\delta) I + \delta (\sqrt{k}, 0,\cdots,0)' (\sqrt{k}, 0,\cdots,0) \\ &= \operatorname{diag} (1+(k-1)\delta , 1-\delta,\cdots,1-\delta) . \end{aligned}$$

Hence Δ is positive definite if and only if $(k-1)^{-1} < \delta < 1$.

2. Derivation of the bound: If $\Sigma \in \mathcal{D}$, then because of symmetry we suspect that the minimizing $B = (1-b)I + be'e \in \mathcal{D}$ and $a = \alpha e$. In any event, we may find the best bound obtainable from such B and α, and an example of sharpness would justify the choice.

By using (1.10) and assuming $a = \alpha e$, $B \in \mathcal{D}$, the bound (1.8) can be written in the form

$$H(\alpha,b) = (1-\alpha)^{-2} \operatorname{tr} B^{-1}(\Sigma + \alpha^2 e'e)$$

$$= (1-\alpha)^{-2} \operatorname{tr} (\Gamma B \Gamma')^{-1} (\Gamma \Sigma \Gamma' + \alpha^2 \Gamma e' e \Gamma') .$$

Since $\Gamma B \Gamma' = \operatorname{diag}(1+(k-1)b, 1-b, \cdots, 1-b)$, $\Gamma \Sigma \Gamma' = \sigma^2 \operatorname{diag}(1+(k-1)\rho, 1-\rho, \cdots, 1-\rho)$, and $e\Gamma' = (\sqrt{k}, 0, \cdots, 0)$, so that

$$H(\alpha,b) = \frac{1}{(1-\alpha)^2}\left[\frac{\sigma^2 + \sigma^2\rho(k-1) + \alpha^2 k}{1+(k-1)b} + \frac{\sigma^2(k-1)(1-\rho)}{1-b}\right] \tag{2.1}$$

$$= \frac{k[\alpha^2 + \sigma^2 + b(\sigma^2 t - \alpha^2)]}{(1-\alpha)^2(1-b)[1+(k-1)b]} ,$$

where

$$t = (k-1)(1-\rho) - 1 . \tag{2.2}$$

The solution of $\partial H/\partial\alpha = 0$ for α is given by

(2.3) $$\alpha_o = -\sigma^2(1+bt)/(1-b) ,$$

from which

(2.4) $$H(\alpha_o,b) = k\,\sigma^2(1+bt)[1+(k-1)b]^{-1}\,[1+\sigma^2-b\,(1-\sigma^2t)]^{-1} .$$

The equation $\partial H(\alpha_o,b)/\partial b = 0$ becomes

(2.5) $$b^2t(1-\sigma^2t) + 2b(1-\sigma^2t) - (\sigma^2+\rho) = 0 ,$$

and has roots

(2.6) $$b = -\frac{1}{t} \pm \frac{\sqrt{1+t\rho}}{t\sqrt{1-\sigma^2t}} = -\frac{1}{t} \pm \frac{\sqrt{(1+t)(k-1-t)}}{t\sqrt{(k-1)(1-\sigma^2t)}} .$$

Using (1.10) , we see that the term $1+t\rho = \{1-\rho\}[1+(k-1)\rho] > 0$ and we assume that $1-\sigma^2t > 0$, so that the roots are real.

B is positive definite if and only if $-(k-1)^{-1} < b < 1$, i.e.

(2.7) $$\frac{k-1-t}{t(k-1)} \leq \pm \frac{\sqrt{(1+t)(k-1-t)}}{t\sqrt{(k-1)(1-\sigma^2t)}} \leq \frac{1+t}{t} .$$

If $t > 0$ $(t < 0)$, then the LHS and RHS are positive (negative),

and the lower sign is impossible. We denote by b_o the root with the positive sign, and by B_o the corresponding matrix.

Lemma: B_o is positive definite if and only if

$$(2.8)\qquad k \geq \sigma^2(k-1)(1+t) .$$

Proof: The condition that B_o be positive definite is equivalent to the inequality (2.7) with the plus sign. The second inequality is (2.8), and the first inequality always holds. ||

An example of sharpness will show that b_o leads to a minimum. We assume (2.8) and evaluate $H(\alpha_o,b_o)$ by showing (using (2.6)) that

$$(2.9)\qquad H(\alpha_o,b_o) = \frac{k\sigma^2(1+b_o t)}{[1+(k-1)b_o][1+\sigma^2-b_o(1-\sigma^2 t)]}$$

$$= \frac{k\sigma^2 t}{[k-2+t\sigma^2+\sigma^2(k-1)]-2b_o(k-1)(1-\sigma^2 t)} .$$

Then substituting for b_o, we obtain

$$(2.10)\qquad H(\alpha_o,b_o) = k\sigma^2 t/(u-2\sqrt{v}) ,$$

where $u = t^2\sigma^2 + t[k-2-(k-1)\sigma^2] + 2(k-1)$, and $v = (1+t)(k-1-t)(k-1)(1-\sigma^2 t)$. Rationalization of the denominator in (2.10) and simplification yields

(2.11) $$H(\alpha_o,b_o) = \frac{k\sigma^2(u+2\sqrt{v})}{[t\sigma^2-k-\sigma^2(k-1)]^2} .$$

After substituting for t from (2.2) we obtain the theorem.

<u>Theorem</u>: Let x be a random vector with $Ex_i=0, Ex_i^2=\sigma^2, Ex_ix_j=\sigma^2\rho\ (i\neq j)$. If (i) $1-\sigma^2t>0$, (ii) $k\geq\sigma^2(k-1)(1+t)$, then

(2.12) $$P \equiv P\{x_1\geq 1 \text{ or } \cdots \text{ or } x_k\geq 1\} \leq H(\alpha_o,b_o)$$

$$= \frac{k\sigma^2\{\sqrt{[1+(k-1)\rho][1+\sigma^2-\sigma^2(k-1)(1-\rho)]} + (k-1)\sqrt{1-\rho}\}^2}{\{k+\sigma^2[1+(k-1)\rho]\}^2} ;$$

otherwise $P\leq 1$.

For the special case $\rho=1$, the inequality reduces to (1.2), and for $\rho=-1/(k-1)$, the bound is $(k-1)\sigma^2$, which reduces to (1.1) for $k=2$.

It should be noted that the bound $H(\alpha_o,b_o)\leq 1$ is equivalent to

$$(\sigma^2(k-1)\sqrt{(1-\rho)[1+(k-1)\rho]} - k\sqrt{1+\sigma^2-\sigma^2(k-1)(1-\rho)})^2 \geq 0$$

3. <u>Sharpness</u>: We show sharpness of (2.12) by exhibiting an example which achieves equality whenever the conditions (i) and (ii) of the theorem are satisfied. For cases that the theorem provides only the trivial bound unity, we give examples when $k=2$.

Let z be a random vector with the following distribution:

$$P\{z = b^{(i)}\} = p/k\ , \qquad i = 1,\cdots,k\ ,$$

$$P\{z = 0\} = 1-p\ ,$$

where $b^{(i)}$ is the i-th row of B_o . If $x = (1-\alpha_o)z + \alpha_o e$ satisfies the conditions of the theorem, then

$$(3.1) \qquad E(z) = -\alpha_o e/(1-\alpha_o) = [1+(k-1)b_o]p\ e/k\ ,$$

$$(3.2) \qquad E(z'z) = (\Sigma + \alpha_o^2\, e'e)/(1-\alpha_o)^2 = p\ B_o^2/k\ .$$

Substituting for α_o from (2.3) in (3.1) and solving for p , we obtain $p = H(\alpha_o, b_o)$, where $H(\alpha_o, b_o)$ is given by (2.9). Because of the special form of Σ , the matrix equation (3.2) is equivalent to the two equations

$$(3.3) \qquad [(1-b_o)^2 + 2b_o(1-b_o) + b_o^2 k]\ p/k = (\sigma^2 + \alpha_o^2)/(1-\alpha_o)^2\ ,$$

$$(3.4) \qquad [2b_o(1-b_o) + b_o^2 k]\ p/k = (\sigma^2\rho + \alpha_o^2)/(1-\alpha_o)^2\ .$$

Substitution of p and α_o from (2.3) in (3.3) and in (3.3) minus (3.4) yields (2.5) with $b = b_o$ in each case. Hence (3.1) and (3.2) are satisfied when $p = H(\alpha_o, b_o)$, that is, when p is given by the bound of (2.12). Since $P\{z_i \geq 1$ for some $i\} = p$, and $z_i \geq 1$ if and only if $x_i \geq 1$, it follows that $x = (1-\alpha_o)z + \alpha_o e$ achieves equality in (2.12).

Now suppose that $k = 2$, in which case conditions (i) and (ii) become $1 + \sigma^2\rho \geq 0$, and $2 \geq \sigma^2(1-\rho)$, respectively.

If $1 + \sigma^2\rho < 0$, then a distribution having the prescribed moments and achieving the bound of one is:

$$P\{(1,-c)\} = P\{(-c,1)\} = p_1/2 ,$$

$$P\{(c,-c)\} = P\{(-c,c)\} = p_2/2 ,$$

$$P\{(1,1)\} = 1 - p_1 - p_2 ,$$

where $p_1 = 2\sigma^2(1+\rho)/(c^2-1)$, $p_2 = (1+\sigma^2\rho)/(1-c^2)$, $c = \frac{1}{2}\left\{\sigma^2(1+\rho)+\sqrt{[\sigma^4(1+\rho)^2+4\sigma^2]}\right\}$. The condition $1 + \sigma^2\rho < 0$ implies that $\sigma^2 > 1$ and $c > 1$. Hence $0 \leq p_1, p_2$, $p_1 + p_2 \leq 1$.

If $2 < \sigma^2(1-\rho)$ and $1 + \sigma^2\rho > 0$, then a distribution with the moments as prescribed in the theorem and achieving the bound of one is:

$$P\{(1,-c)\} = P\{(-c,1)\} = p/2 , \quad P\{(d,d)\} = 1 - p ,$$

where

$$p = \frac{2d}{2d+c-1} = \frac{2\sigma^2(1-\rho)}{(1+c)^2} , \quad c = -\frac{1+\sqrt{(1-\rho^2)(1+\sigma^2\rho)}}{\rho} ,$$

($c = \sigma^2/2$ if $\rho = 0$). The condition $2 < \sigma^2(1-\rho)$ implies that $c > 1$, which in turn implies that $p < 1$. It also implies $1 + c < \sigma^2(1-\rho)$ which is equivalent to $d = p(c-1)/2(1-p) > 1$.

If $1 + \sigma^2\rho = 0$, then the above distribution with $d = 1$, $c = \sigma^2$, $p = 2/(1+\sigma^2)$ is the required example.

4. An inequality involving variances. If $x_1, \cdots, x_k$ are random variables with $Ex_i = 0$, $Ex_i^2 = \sigma_i^2$, $i = 1, \cdots, k$, then by (1.1)

$$(4.1)\qquad P\{|x_1| \geq 1 \text{ or } \cdots \text{ or } |x_k| \geq 1\} \leq \sum_1^k P\{|x_j| \geq 1\} \leq \sum_1^k \sigma_j^2 .$$

This inequality was proved to be sharp in [1], and the unique distribution attaining equality has zero covariances.

The corresponding one-sided inequality, by (1.2), is

$$(4.2)\qquad P\{x_1 \geq 1 \text{ or } \cdots \text{ or } x_k \geq 1\} \leq \sum_1^k P\{x_j \geq 1\} \leq \sum_1^k \sigma_j^2/(1+\sigma_j^2) .$$

If the bound is ≤ 1, the unique distribution attaining equality is

$$P\{(1, -\sigma_2^2, -\sigma_3^2, \cdots, -\sigma_k^2)\} = \sigma_1^2/(1+\sigma_1^2) ,$$

$$P\{(-\sigma_1^2, 1, -\sigma_3^2, \cdots, -\sigma_k^2)\} = \sigma_2^2/(1+\sigma_2^2) ,$$

$$\cdots$$

$$P\{(-\sigma_1^2, -\sigma_2^2, \cdots, -\sigma_{k-1}^2, 1)\} = \sigma_k^2/(1+\sigma_k^2) ,$$

$$P\{(-\sigma_1^2, -\sigma_2^2, \cdots, -\sigma_{k-1}^2, -\sigma_k^2)\} = 1 - \sum_1^k \sigma_j^2/(1+\sigma_j^2) .$$

Uniqueness follows by an argument similar to that used in [1]. We note that in this case the covariances $Ex_i x_j = -\sigma_i^2 \sigma_j^2$ are not zero.

An alternative proof of (4.2) following the procedures of Section 1 is to choose $B = I$ in (1.8) and minimize $\operatorname{tr} D_{1-a}^{-2}(\Sigma + a'a)$ with respect to $a < 1$. The minimizing $a_j = -\sigma_j^2$.

REFERENCES

[1] I. Olkin and J. W. Pratt, "A multivariate Tchebycheff inequality," Ann. Math. Stat., Vol. 29 (1958), pp. 226-234.

[2] J. V. Uspensky, Introduction to Mathematical Probability, McGraw Hill, New York, 1937.

A ONE-SIDED ANALOG OF KOLMOGOROV'S INEQUALITY

by

Albert W. Marshall

1. Introduction and summary.

It is well-known (see e.g. [1] p. 198) that for every positive ϵ and every square integrable random variable X with zero expectation,

$$P\{X \geq \epsilon\} \leq E(X^2)/[\epsilon^2 + E(X^2)] .$$

In this paper an inequality is obtained that generalizes this in the same way that Kolmogorov's inequality generalizes Chebyshev's inequality. The inequality is proved in Section 2 and an example is given to show that equality can be achieved. The possibility of obtaining greater generality is also discussed. In Section 3 an extension to continuous parameter martingales is obtained, and a condition under which equality can be achieved is given.

2. The inequality.

Theorem 2.1: Let $X_1, X_2, \ldots, X_n$ be random variables with $E(X_1) = 0$, $E(X_i \mid X_1, X_2, \ldots, X_{i-1}) = 0$ a.e., $i = 2, 3, \ldots, n$, and $E(X_i^2) = \sigma_i^2 < \infty$, $i = 1, 2, \ldots, n$. Then for every positive ϵ,

(2.1) $$P\{\max_{1 \leq i \leq n}(X_1 + X_2 + \cdots + X_i) \geq \epsilon\} \leq \text{[illegible]} \epsilon^2 + \text{[illegible]} .$$

where $s_n = \sum_{i=1}^{n} \sigma_i^2$.

Proof: Let $F(x) = F(x_1,x_2,\cdots,x_n) = \dfrac{(\epsilon \sum_{i=1}^{n} x_i + s_n)^2}{(\epsilon^2 + s_n)^2}$, and let

$A_k = \{X_1 + X_2 + \cdots + X_i < \epsilon ,\ i = 1,2,\cdots,k-1 ,\ X_1 + X_2 + \cdots + X_k \geq \epsilon\}$

$k = 1,2,\cdots,n$. Then

$$\int F(X)dP \geq \sum_{k=1}^{n} \int_{A_k} F(X)dP =$$

$$= \frac{1}{(\epsilon^2 + s_n)^2} \sum_{k=1}^{n} \int_{A_k} [(\epsilon \sum_{i=1}^{k} X_i + s_n)^2 + 2(\epsilon \sum_{i=1}^{k} X_i + s_n)(\epsilon \sum_{i=k+1}^{n} X_i) + (\epsilon \sum_{i=k+1}^{n} X_i)^2]dP$$

$$\geq \frac{1}{(\epsilon^2 + s_n)^2} \sum_{k=1}^{n} \int_{A_k} (\epsilon \sum_{i=1}^{k} X_i + s_n)^2 dP \geq \sum_{k=1}^{n} P(A_k) = P\{\max_{1 \leq i \leq n} (X_1 + \cdots + X_i) \geq \epsilon\} .$$

Since $\int F(X)dP = s_n/(\epsilon^2 + s_n)$, the proof is complete..

To show that equality can be achieved in (2.1), let $s_k = \sum_{i=1}^{k} \sigma_i^2$,

$k = 1,2,\cdots,n$, and let $Z = (Z_1,Z_2,\cdots,Z_n)$ be a random variable having the following distribution:

$$P\{Z = (\epsilon,0,\cdots,0)\} = \frac{\sigma_1^2}{\epsilon^2 + s_1} = p_1 ,$$

$$P\{Z = \tfrac{1}{\epsilon}(-\sigma_1^2,-\sigma_2^2,\cdots,-\sigma_k^2,\epsilon^2+s_{k-1},0,\cdots,0)\} = \frac{\epsilon^2 \sigma_k^2}{(\epsilon^2+s_{k-1})(\epsilon^2+s_k)} = p_k,\ k = 2,3,\cdots,n$$

$$P\{Z = \tfrac{1}{\epsilon}(-\sigma_1^2,-\sigma_2^2,\cdots,-\sigma_n^2)\} = \frac{\epsilon^2}{\epsilon^2 + s_n} .$$

It is easily verified (by induction on j) that

(2.2) $$\sum_{k=1}^{j} p_k = 1 - \epsilon^2/(\epsilon^2 + s_j) \ , \quad j = 1,2,\cdots,n \ ,$$

so that this is a valid probability distribution. Let us check that the random variables $Z_1, Z_2, \cdots, Z_n$ satisfy the conditions of Theorem 2.1 . Trivially $E(Z_1) = 0$. For $1 < j \leq n$, $E(Z_j | Z_{j-1} \neq -\sigma_{j-1}^2/\epsilon) = 0$ a.e., since $Z_{j-1} \neq -\sigma_{j-1}^2/\epsilon$ implies $Z_j = 0$ with probability one. Using (2.2), it follows that $E(Z_j | Z_{j-1} = -\sigma_j^2/\epsilon)$ is a.e. proportional to $p_j(\epsilon^2 + s_{j-1})/\epsilon - (1 - \sum_{k=1}^{j} p_k)\sigma_j^2/\epsilon = 0$. Hence $E(Z_j | Z_{j-1}) = 0$ a.e., and it is clear that $E(Z_j | Z_{j-1}) = E(Z_j | Z_1, Z_2, \cdots, Z_{j-1})$ a.e. Again using (2.2), it follows that $E(Z_j^2) = p_j(\epsilon^2 + s_{j-1})^2/\epsilon^2 + (1 - \sum_{k=1}^{j} p_k)\sigma_j^4/\epsilon^2 = \sigma_j^2$, $j = 1,2,\cdots,n$. Thus the random variable Z satisfies the conditions of Theorem 2.2; furthermore, equality holds in (1.1) whenever $X = (X_1, X_2, \cdots, X_n) = Z$ a.e.

It is natural now to ask what the best upper bound is for $P(\max_{1 \leq i \leq n} \frac{1}{\epsilon_i}(X_1 + X_2 + \cdots + X_i) \geq 1)$ under the conditions of Theorem 2.1 (where $\epsilon_i > 0$, $i = 1,2,\cdots,n$) . Unfortunately this bound has no simple expression even for small n , and is not easily obtained. It is given here only for $n = 2$.

Theorem 2.3: Let X_1 and X_2 be random variables with $E(X_1) = 0$, $E(X_2 | X_1) = 0$ a.e., and $E(X_i^2) = \sigma_i^2 < \infty$, $i = 1,2$. Then if $\epsilon_1 > 0$ and $\epsilon_2 > 0$,

$$(2.3)\qquad P\{X_1 \geq \epsilon_1 \text{ or } X_1 + X_2 \geq \epsilon_2\} \leq \frac{\sigma_2^2 + \sigma_1^2(\alpha_2/\alpha_1)^2}{\sigma_2^2 + \sigma_2^2/\alpha_1} ,$$

where $\alpha_i = \sigma_i^2 + \eta_1\eta_i$, $i = 1,2$, and $\eta_1 = \min(\epsilon_1,\epsilon_2)$, $\eta_2 = \epsilon_2$.

Proof: Let $F(x_1,x_2) = c_1 F_1^2(x_1) + c_2 F_2^2(x_1 + x_2)$ where

$$c_1 = \frac{\eta_1^2}{\alpha_1^2} - \frac{\eta_1^2(\alpha_2+\sigma_2^2)^2}{(\alpha_2^2+\sigma_2^2\alpha_1)^2} , \quad c_2 = \frac{\eta_1^2\sigma_2^2}{(\alpha_2^2+\sigma_2^2\alpha_1)^2} ,$$

$$F_1(x) = (x + \frac{\sigma_1^2}{\eta_1}) , \qquad F_2(x) = (x + \frac{\sigma_1^2}{\eta_1} + \frac{\sigma_2^2\alpha_1}{\eta_1\alpha_2}) ,$$

and let $A_1 = \{X_1 \geq \eta_1\}$, $A_2 = \{X_1 < \eta_1 , X_1 + X_2 \geq \eta_2\}$. Since $\alpha_2 \geq \alpha_1 > 0$, it follows that $c_1 \geq 0$, and

$$\int_{A_1} F(X_1,X_2)dP = \int_{A_1} [c_1F_1^2(X_1) + c_2F_2^2(X_1) + c_2X_2F_2(X_1) + c_2X_2^2]dP$$

$$\geq \int_{A_1} [c_1F_1^2(X_1)+c_2F_2^2(X_1)]dP \geq \int_{A_1} [c_1F_1^2(\eta_1)+c_2F_2^2(\eta_1)]dP = P(A_1) ,$$

$$\int_{A_2} F(X_1,X_2)dP \geq \int_{A_2} c_2F_2^2(X_1 + X_2)dP \geq \int_{A_2} c_2F_2^2(\eta_2)dP = P(A_2) .$$

Thus $\int F(X_1,X_2)dP \geq P(A_1)+P(A_2) = P\{X_1 \geq \eta_1$ or $X_1 + X_2 \geq \eta_2\}$ $\geq P\{X_1 \geq \epsilon_1$ or $X_1 + X_2 \geq \epsilon_2\}$. It is straightforward to verify that

upon integrating the function $F(X_1,X_2)$, one obtains the bound given in (2.3), and this completes the proof.

To show that equality can be achieved in (2.3), let (Z_1,Z_2) have the following joint distribution:

$$P\{(Z_1,Z_2) = (\eta_1,0)\} = \sigma_1^2/\alpha_1$$

$$P\{(Z_1,Z_2) = (-\frac{\sigma_1^2}{\eta_1}, \frac{\alpha_2}{\eta_1})\} = \frac{\eta_1^2\,\sigma_2^2}{\alpha_1\sigma_2^2+\alpha_2^2},$$

$$P\{(Z_1,Z_2) = (-\frac{\sigma_1^2}{\eta_1}, -\frac{\sigma_2^2\,\alpha_1}{\eta_1\,\alpha_2})\} = \frac{\eta_1^2\,\alpha_2^2}{\alpha_1(\alpha_1\sigma_2^2+\alpha_2^2)}.$$

It is easily verified that (Z_1,Z_2) satisfies the conditions of Theorem 2.2 . Furthermore,

$$P\{(Z_1,Z_2) = (\eta_1,0)\} + P\{(Z_1,Z_2) = (-\frac{\sigma_1^2}{\eta_1}, \frac{\alpha_2}{\eta_1})\} = \frac{\sigma_2^2+\sigma_1^2(\alpha_2/\alpha_1)^2}{\sigma_2^2+(\alpha_2^2/\alpha_1)}$$

$$= P\{Z_1 \geq \eta_1 \text{ or } Z_1+Z_2 \geq \eta_2\} = P\{Z_1 \geq \epsilon_1 \text{ or } Z_1+Z_2 \geq \epsilon_1+\epsilon_2\},$$

so that equality holds in (2.3) whenever $(X_1,X_2) = (Z_1,Z_2)$ a.e.

Several inequalities follow from (2.3) simply by a change of variables. The corollaries below are given to illustrate the possibilities.

<u>Corollary 2.3</u>: Let X_1 and X_2 be random variables with $E(X_1) = a$,

$E(X_2|X_1) = bX_1 + c$ a.e.(where $b \neq -1$), and $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$, $i = 1,2$. Then if $\epsilon_1 - a > 0$ and $[\epsilon_2 - \delta(a+ab+c)]/|b+1| > 0$ where $\delta = \mathrm{sign}(b+1)$,

$$(2.4) \qquad P\{X_1 \geq \epsilon_1 \text{ or } \delta(X_o + X_2) \geq \epsilon_2\} \leq \frac{\sigma_2^2 - b^2\sigma_1^2 + \sigma_1^2\{(b+1)\alpha_2/\alpha_1\}^2}{\sigma_2^2 - b^2\sigma_1^2 + [(b+1)^2\alpha_2^2/\alpha_1]}$$

where $\alpha_i = \sigma_i^2 + \eta_1\eta_i$, $i = 1,2$, and $\eta_2 = [\epsilon_2 - \delta(a+ab+c)]/|b+1|$, $\eta_1 = \min(\epsilon_1 - a, \eta_2)$.

<u>Proof</u>: This follows from Theorem 2.2 by making the change of variables

$$X_1' = X_1 + a, \quad X_2' = bX_1 + (b+1)X_2 + ab+c, \quad \epsilon_1' = \epsilon_1 + a, \quad \epsilon_2' = \epsilon_2|b+1| + \delta(a+ab+c)$$

and dropping the primes.

Note that by taking $a = b = c = 0$ in this corollary, one obtains Theorem 2.2.

<u>Corollary 2.4</u>: Let X_1 and X_2 be random variables such that $E(X_i) = \mu_i$, $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$, $i = 1,2$, $\mathrm{Cov}(X_1,X_2) = \sigma_{12} \neq 0$, and suppose that the regression of X_2 on X_1 is linear. Then if $\epsilon_1 - \mu_1 > 0$ and $(\delta\epsilon_2 - \sigma_1^2\mu_2)/\sigma_{12} > 0$ where $\delta = \mathrm{sign}\,\sigma_{12}$,

$$(2.5) \qquad P(X_1 \geq \epsilon_1 \text{ or } \delta X_2 \geq \epsilon_2) \leq \frac{\sigma_1^2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2) + \sigma_1^2(\alpha_2\sigma_{12}/\alpha_1)^2}{\sigma_1^2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2) + (\alpha_2^2\sigma_{12}^2/\alpha_1)},$$

where $\alpha_i = \sigma_i^2 + \eta_i\eta_i$, $i = 1,2$, and $\eta_2 = (\delta\epsilon_2 - \sigma_1^2\mu_2)/\sigma_{12}$,
$\eta_1 = \min(\epsilon_1 - \mu_1, \eta_2)$.

Proof: To obtain (2.5) from (2.3), make the change of variables

$$X_1' = X_1 + \mu_1, \quad X_2' = [\sigma_{12}'(X_1 + X_2)/\sigma_1'^2] + \mu_2, \quad \epsilon_1' = \epsilon_1 + \mu_1 \quad \text{and} \quad \epsilon_2' = \delta(\epsilon_2\sigma_{12}' + \sigma_1'^2\mu_2)$$

in (2.3) and then remove the primes.

3. An extension to continuous parameter martingales.

We begin by assuming that the underlying probability space is such that P is complete. Then we have the following

Theorem 3.1: Let $\{Y_t, t \geq 0\}$ be a separable martingale such that for all $t \geq 0$, $E(Y_t) = 0$ and $E(Y_t^2) = \sigma^2(t) < \infty$. Then for every positive ϵ and τ,

$$P\left\{\sup_{t\in[0,\tau]} Y_t \geq \epsilon\right\} \leq \frac{\sigma^2(\tau)}{\epsilon^2 + \sigma^2(\tau)} . \tag{3.1}$$

Proof: Let $0 = t_1 \leq t_2 \leq \cdots \leq t_n = \tau$. Then $X_1 = Y_{t_1}$ and $X_i = Y_{t_i} - Y_{t_{i-1}}$, $i = 2,3,\cdots,n$ satisfy the conditions of Theorem 2.1, so that

$$P\left\{\max_{1 \leq i \leq n} Y_{t_i} \geq \epsilon\right\} \leq \sigma^2(\tau)/[\epsilon^2 + \sigma^2(\tau)] . \tag{3.2}$$

Let S be a countable set satisfying the definition of separability and containing the points 0 and τ . Taking the supremum of the left side of (3.2) over all finite subsets of $S \cap [0,\tau]$, we obtain

$$P\left\{\sup_{t \in S \cap [0,\tau]} Y_t \geq \epsilon\right\} \leq \sigma^2(\tau)/[\epsilon^2 + \sigma^2(\tau)] .$$

But

$$P\left\{\sup_{t \in S \cap [0,\tau]} Y_t \geq \epsilon\right\} = P\left\{\sup_{t \in [0,\tau]} Y_t \geq \epsilon\right\} ,$$

and the proof is complete.

Theorem 3.2: Equality can be achieved in (3.1) if $\sigma^2(\cdot)$ is right continuous.

Proof: In order to define a martingale that achieves equality in (3.1), let $\Omega = \{-1\} \cup [0,\infty)$, $\mathcal{B}$ be the Borel subsets of Ω , and let P be the probability measure defined on $\mathcal{B}$ by

$$P(B) = (\epsilon^2/[\epsilon^2 + \lim_{x \to \infty} \sigma^2(x)]) \chi_{B \cap \{-1\}} + \mu(B \cap [0,\infty)) ,$$

where χ_E is the characteristic function of the set E and μ is the measure induced on the Borel subsets of $[0,\infty)$ by the right continuous distribution function $\sigma^2(\cdot)/[\epsilon^2 + \sigma^2(\cdot)]$. Let $\{Z_t, t \geq 0\}$ be defined on $(\Omega, \mathcal{B}, P)$ by

$$Z_t(\omega) = \begin{cases} -\sigma^2(t)/\epsilon & 0 \le t < \omega \\ \epsilon & 0 \le \omega \le t \\ 0 & \omega = -1 . \end{cases}$$

Then

$$P\left\{\sup_{t\in[0,\tau]} Z_t \ge \epsilon\right\} = P\left\{0 \le \omega \le \tau\right\} = \sigma^2(\tau)/[\epsilon^2 + \sigma^2(\tau)] ,$$

and it remains only to verify that the process $\{Z_t, t \ge 0\}$ satisfies the conditions of Theorem 3.1. We compute

$$E(Z_t) = [-\sigma^2(t)P(\omega > t)/\epsilon] + \epsilon P(0 \le \omega \le t) = 0 ,$$

and similarly obtain $E(Z_t^2) = \sigma^2(t)$, $t > 0$. Clearly $E(Z_t|Z_s = \epsilon) = Z_s$ where $0 \le s < t$ are fixed. Let $\theta = E(Z_t|Z_s = -\sigma^2(s)/\epsilon)$; using the relation

$$0 = E(Z_t) = E[E(Z_t|Z_s)] = \epsilon P(Z_s = \epsilon) + \theta P(Z_s = -\sigma^2(s)/\epsilon) ,$$

we obtain $\theta = -\sigma^2(s)/\epsilon$. Hence the process $\{Z_t, t \ge 0\}$ is a martingale satisfying the conditions of Theorem 3.1 and achieving equality in (3.1).

Reference

[1] J.V. Uspensky, Introduction to Mathematical Probability, Mc Graw-Hill, New York, 1937 .